Computers and Electronics in Agriculture 68 (2009) 52–61

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Artificial neural networks for automated year-round temperature prediction

Brian A. Smith [a,b], Gerrit Hoogenboom [a,c,*], Ronald W. McClendon [a,d]

[a] Institute for Artificial Intelligence, The University of Georgia, Athens, GA 30602, USA
[b] Computer Science Department, The University of Georgia, Athens, GA 30602, USA
[c] Department of Biological and Agricultural Engineering, The University of Georgia, Griffin, GA 30223, USA
[d] Department of Biological and Agricultural Engineering, Driftmier Engineering Center, The University of Georgia, Athens, GA 30602, USA

ARTICLE INFO

Article history:
Received 31 March 2008
Received in revised form 12 February 2009
Accepted 22 April 2009

Keywords:
Artificial intelligence
Neural network
Temperature prediction
Frost protection
Fruit crops
Vegetable crops

ABSTRACT

Crops and livestock in most of the southeastern United States are susceptible to potential losses due to extreme cold and heat. However, given suitable warning, agricultural and horticultural producers can mitigate the damage of extreme temperature events. To provide such a warning, air temperature prediction models are needed at horizons ranging from 1 to 12 h. The goal of this project was to explore the application of artificial neural networks (ANNs) for the prediction of air temperature during the entire year based on near real-time data. Ward-style ANNs were developed using detailed weather data collected by the Georgia Automated Environmental Monitoring Network (AEMN). The ANNs were able to provide predictions throughout the year, with a mean absolute error (MAE) of the year-round models that was less during the winter months than the MAE of the models resulting from the application of previously developed winter-specific models. The prediction MAE for a year-round evaluation set ranged from 0.516 °C at the one-hour horizon to 1.873 °C at the twelve-hour horizon. A detailed graphical analysis of MAE by time-of-year and time-of-day was also performed. A tendency to over-predict temperatures during summer afternoons was associated with localized cloud cover during that period. The inclusion of rainfall as input to the model was also shown to improve prediction accuracy. In addition, two simple ensemble techniques were explored and neither parallel nor series aggregation was found to reduce prediction errors. When simulated over two extreme temperature events, the models were capable of rapidly adjusting predictions on the basis of new information. The final models were applied to prediction horizons of 1–12 h and deployed on the website of the Georgia AEMN (www.GeorgiaWeather.net) for use as a general, year-round decision support tool.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

Damage to plants and animals caused by extreme temperatures is a serious concern for agricultural producers. In the state of Georgia and elsewhere in the southeastern United States, frost damage is a problem during the late winter and early spring when bud formation and flowering occur in many horticultural crops. Late freezes can damage floral buds of fruit, leading to reduced harvests. Orchard heaters, irrigation, wind machines and other devices can protect fruit trees and bushes from severe frost damage, provided that growers are given adequate warning of freezing conditions. Extreme heat cannot only damage plants, but can cause heat exhaustion in both livestock and agricultural workers (National Agricultural Statistical Service, 2005). The accurate prediction of extreme temperature events could also be valuable in decision support systems for energy management and transportation. To be suitable for decision support, temperature prediction models must be accurate, robust, and available 24 h a day.

Artificial neural networks (ANNs) have been used in a number of prediction studies that involve atmospheric time series data. Yi and Prybutok (1996) predicted daily maximum ozone levels in Texas metropolitan areas with a standard three-layer ANN model with nine inputs and four hidden nodes and found it to be superior to statistical methods. A three-layer ANN model with 17 inputs was developed by Jiang et al. (2004) to predict the air pollution levels of cities in China. Inputs to the models were not site-specific, allowing the model to be applied to a number of locations across China. Air temperature, wind speed, and relative humidity in Saskatchewan, Canada were predicted 24 h in advance by ANN models developed and applied by Maqsood et al. (2004). They found that combining the outputs of a standard feed-forward ANN, a recurrent ANN, a radial basis function network, and a Hopfield network into a simple “winner-take-all” ensemble led to more accurate predictions of wind speed, relative humidity, and air temperature than any

* Corresponding author at: Department of Biological and Agricultural Engineering, The University of Georgia, 1109 Experiment Street, Griffin, GA 30223, USA. Tel.: +1 770 228 7216; fax: +1 770 228 7218.
E-mail address: Gerrit@uga.edu (G. Hoogenboom).

0168-1699/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.compag.2009.04.003

of the individual component networks. Ruano et al. (2006) used a multi-objective genetic algorithm to develop a radial basis function ANN model for the prediction of air temperature in a secondary school building in Portugal. Air-conditioning control scheme simulations indicated that temperatures could be more consistently managed and that air conditioner run times could be reduced using an ANN model. Ghielmi and Eccel (2006) compared the use of ANNs with traditional mathematical models to predict minimum temperatures. They found that the ANNs provided comparable accuracy and offered the advantage of including inputs not used in traditional equations. The authors also suggested that predictions of minimum temperatures should be included on a website for dissemination to users.

In 1991, the University of Georgia initiated the Georgia Automated Environmental Monitoring Network (AEMN) to collect weather data from sites across the state of Georgia (Hoogenboom, 2000). This network has expanded to more than 75 sites that gather local information for a range of environmental variables. The data are obtained at 1 s intervals and aggregated into 15 min summaries. Each site consists of an automated, solar-powered station that periodically downloads the data to a central server located at the University of Georgia Griffin Campus. The weather data are gathered primarily from rural locations for which detailed observations from the National Weather Service are unavailable. The data are disseminated via the AEMN website www.GeorgiaWeather.net. The website also provides a number of calculators, maps, and decision support tools to assist users. Twelve-hour air temperature predictors are available during the winter and early spring. Accurate temperature predictions during this time of year can provide advance warning of upcoming freezes, allowing horticultural producers to employ mitigation techniques such as irrigation or orchard heating to reduce the potential damage to the developing flowers and fruits.

The data provided by the AEMN have been used to develop ANN models for the prediction or estimation of atmospheric variables. Weather observations from the Tifton, Georgia along with three out-of-state sites were used to develop ANNs to estimate daily solar radiation based on daily minimum and maximum air temperatures, rainfall, and calculated values of clear-sky radiation and length of day (Elizondo et al., 1994). Data from three AEMN sites were used by Bruton et al. (2000) to develop ANN-based models to estimate daily pan evaporation. The predictions of the ANN model were a slight improvement over those of statistical regression models. Li et al. (2004) developed ANNs to estimate daily maximum and minimum air temperature as well as total solar radiation for sites in Tifton and Griffin, Georgia using AEMN data from nearby sites.

Several studies have used AEMN data to develop ANN models as an aid for frost protection decision support. Air temperature, solar radiation, wind speed, and relative humidity were found to be useful meteorological inputs for air temperature prediction ANNs during the winter and early spring (Jain et al., 2003, 2006). Inputs to the networks, which predicted air temperature from 1 to 12 h ahead, included up to 6 h of prior observations for each input series. In addition, the work encoded the time of day at the point of prediction using four cyclic input variables. Ramyaa (2004) developed classification ANNs to predict freeze events rather than using air temperature for prediction. The networks were trained to classify an upcoming twelve-hour period as a freeze, near-freeze, or non-freeze event. In addition to the inputs identified by Jain et al. (2006), these classification networks included current and prior rainfall observations. Ward-style ANN models to predict dew point temperatures up to 12 h in advance were developed by Shank et al. (2008a). Ensemble ANNs were subsequently used to improve the accuracy of the dew point temperature predictions (Shank et al., 2008b). These ANNs were implemented on the AEMN website www.GeorgiaWeather.net. Predictions of dew point temperature can help assess the severity of frost and freeze events when coupled with accurate air temperature predictions.

Smith et al. (2006) found that training multiple instances of a model, each with different initial randomly-assigned weights, increased the likelihood that ANN design alternatives were evaluated accurately. This approach was applied to the development of winter-only air temperature prediction models with the goal of developing models which would be more accurate than those of Jain et al. (2003). The models were developed with prediction horizons from 1 to 12 h ahead using the Ward-style ANN architecture and AEMN weather data. Observation patterns from January through early April were used for ANN development and evaluation. Rainfall observations were included as inputs based on the results of the freeze classification approach of Ramyaa (2004). Smith et al. (2006) found that models which included cyclic, day-of-year variables and 24 h of prior data as inputs produced more accurate predictions than models without such inputs. The inclusion of day-of-year variables suggested the possibility of extending the methodology to a year-round prediction scheme. Prabha and Hoogenboom (2008) evaluated the use of the Advanced Research Weather Research and Forecasting model for two episodic frost events in the southeastern U.S. The accuracy of the minimum temperature prediction ranged from 90% for the April event to 80% for the December event. No attempt to include these simulations in a real-time mode was pursued due to the excessive computational requirements.

The goal of the research presented herein was to develop a set of year-round ANN models for air temperature prediction for 1–12 h ahead for inclusion in general, year-round, real-time decision support aids. The objectives related to this goal were: (1) a comparison of the temperature prediction accuracy of year-round models to that of existing winter models, (2) a determination of the effect on accuracy of including rainfall input terms, (3) an examination of ensemble network techniques, and (4) an analysis of prediction errors based on the day-of-year and time-of-day data to assess the suitability of the models for general decision-making.

2. Methodology

2.1. Data sets

Meteorological data observations from the AEMN system were formatted and scaled into input-target patterns. The patterns were partitioned into the following data sets: a development set, a selection set, and an evaluation set. The development set was used to train multiple networks for each model and to choose a preferred network from among those instantiating the same model. Distinct instantiations of the same model differed in the initial random weights assigned to the network and the order of presentation of training patterns. The development set was created from patterns recorded during the years 1997 to 2000 and consisted of approximately 1.25 million patterns. The patterns in the development set were drawn from the weather stations located in Alma, Arlington, Attapulgus, Blairsville, Fort Valley, Griffin, Midville, Plains, and Savannah. These sites are located across the state and represent both geographical and agricultural diversity. For example, Arlington and Attapulgus are in the Georgia Coastal Plain, Blairsville is located in northern Georgia's Appalachian Mountains, Griffin is in the central Piedmont, and Savannah is located on the Atlantic coast. Patterns from Brunswick, Byron, Cairo, Camilla, Cordele, Dearing, Dixie, Dublin, Homerville, Nahunta, Newton, Valdosta, and Vidalia were used for model selection and evaluation. Collectively, these stations represent important agricultural production areas in the southern and central parts of the state of Georgia. The selection set patterns were drawn from 2001 to 2003 and numbered approximately 1.25 million. The MAE for this set was used to select between competing ANN design alternatives. The evaluation set from the years 2004 to 2005 consisted of approximately 800,000 patterns. This data set was used to calculate an independent error measure for the selected model. The sites used in the model development and evaluation sets were the same as those used by Smith et al. (2006).

Previous models developed by Jain et al. (2006) and Smith et al. (2006) predicted air temperature to aid in frost prediction and were trained on patterns generated from observations that occurred during the first 100 days of the year. The patterns used were also restricted to those in which the temperature at the time of the prediction was less than or equal to 20 °C. As a result of these constraints, the models would not be applicable for higher temperatures or during other times of the year.

The goal of the current research was to develop models that perform well over the entire year without sacrificing accuracy over winter observations. Therefore, winter-period subsets of the development, selection, and evaluation sets were created. Decisions regarding preferred networks and models were based on the MAE for year-round data sets, but the MAE calculated using the winter subsets was considered as an indication of model accuracy during the winter months.

Current values and 24 h duration of prior observations for air temperature, solar radiation, wind speed, rainfall, and relative humidity from the time of prediction were used as inputs for the ANN models. The hourly rates of change in each of the five weather variables at the prediction point and at 1-h intervals over the previous day were also included as inputs for the models. These 250 inputs were rescaled using a linear transformation such that all observations used in model development were in the range [0.1,0.9]. The scaled value, x_{scaled}, of an observation x was determined by

$$x_{\text{scaled}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} 0.8 + 0.1 \quad (1)$$

where $x_{\min}$ and $x_{\max}$ were the minimum and maximum values of the variable found in the development data set. Because the output of the network was restricted to the domain [0,1], it was necessary to map the output signal back to the range of expected temperatures. The inverse of the linear function used to scale the input temperature was used for this purpose.

The time-of-day and the day-of-year data were each encoded as four cyclic variables using triangular fuzzy logic-type membership functions in the range [0,1]. The four time-of-day membership functions are presented in Fig. 1. The variable midnight "wraps around" the day, with a maximum value at 2400 h. The figure is a smooth, continuous representation of the membership functions. In practice, AEMN observations are aggregated at 15 min intervals beginning at the top of the hour. The values of the time-of-day membership functions were determined by the hour in which the observation occurred. Day-of-year variables were treated in a similar manner, using four membership functions to represent seasonality. With the inclusion of these eight cyclic variables, each model had a total of 258 inputs.

2.2. Model development

Ward-style (Ward Systems Group, Inc. NeuroShell 2, Frederick, MD (http://www.wardsystems.com/)) ANNs with a prediction horizon of 4 h were used for model development, similar to previous studies of air temperature predictions with AEMN data (Jain et al., 2003; Ramyaa, 2004; Smith et al., 2006). Design decisions based upon this horizon were later used to develop models to predict from 1 to 12 h ahead. The Ward-style ANNs are feed-forward, backpropagation networks that implement multiple activation functions in a single hidden layer. In the Ward-style networks used in this study,

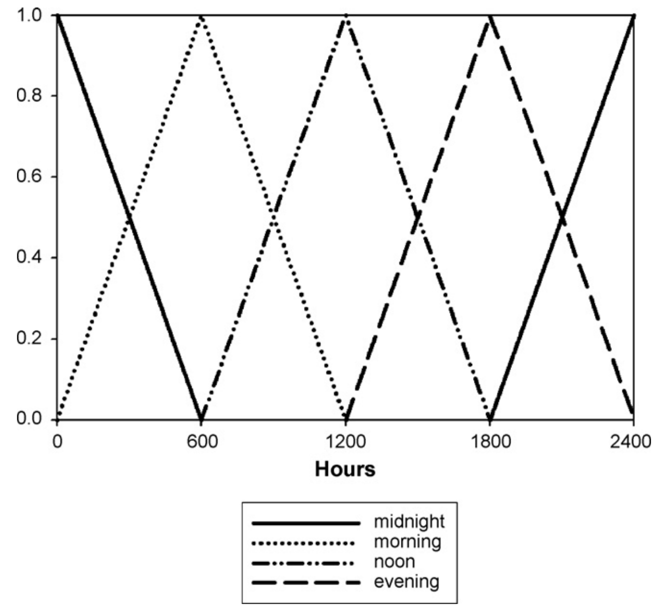

Fig. 1. Triangular fuzzy logic-type membership functions for time-of-day inputs to the models.

the signal of the kth output node, z_k, is given by

$$z_k = g\left(\sum_{j=0}^{J} \beta_{kj} y_i\right), \quad (2)$$

where y_j is the output of the jth hidden node in the network, β_{kj} is an adjustable weight parameter, and

$$g(n) = \frac{1}{1 + \exp(-n)}. \quad (3)$$

The term y_0 is set to a constant value of 1 and the coefficient β_{k0} is the bias of the kth output node. The value of the summation over the J hidden nodes is referred to as the induced local field of the node (Haykin, 1999). Some ANN architectures use the same activation function, g, for each node in the hidden layer. The Ward-style architecture, in contrast, makes use of multiple activation functions in a single hidden layer. When x_i is the output of the ith input and α_{ji} is a weighting factor, the value of y_j, the jth hidden node in a Ward-style network, is given by

$$y_j = f_j\left(\sum_{i=0}^{I} \alpha_{ji} x_i\right), \quad (4)$$

where

$$0 < j_1 < j_2 < J \quad (5)$$

and

$$f_j(n) = \begin{cases} \tan h(n), & \text{for } 0 < j \le j_1 \\ \exp(-n^2), & \text{for } j_1 < j \le j_2 \\ 1 - \exp(-n^2), & \text{for } j_2 < j \le J \end{cases} \quad (6)$$

As in Eq. (2), $x_0 = 1$, so that each α_{j0} serves as a bias for its respective hidden node. All "slabs" of nodes with the same activation function were constrained to be of an equal size. As the ANNs were trained to predict temperature at a single prediction horizon, there was only one output node z.

The α and β terms in Eqs. (2) and (4) are adjustable weight coefficients. Training a neural network via error back propagation (EBP) is an attempt to identify a set of weights that reduces the mean squared error of the output node over a development set. EBP performs a local gradient search in the space of possible weights. The algorithm was applied after the presentation of each training pattern and a single feed-forward pass followed by a weight adjustment constituted a learning event. For a training pattern n with target $t(n)$, the prediction error and error energy of a network with output $z(n)$, are $e(n)$ and $E(n)$, respectively, where

$$e(n) = t(n) - z(n) \tag{7}$$

and

$$E(n) = \frac{1}{2}e(n)^2. \tag{8}$$

The change to a weight w in light of a training pattern n and a learning rate η was given by the delta rule:

$$\Delta w = -\eta \frac{\partial E(n)}{\partial w}. \tag{9}$$

The value of the last term in Eq. (9) is the partial derivative of the network error energy with respect to a selected weight w. The calculation of this term depends on whether the weight in question feeds a signal to the output layer or the hidden layer. In the former case, w is β_{kj}. Differentiating Eq. (2) yields

$$\Delta\beta_{kj} = \eta e(n) g' \left(\sum_{j=0}^{J} \beta_{kj} y_j \right) y_j. \tag{10}$$

For weights leading into the hidden layer

$$\Delta\alpha_{ji} = \eta e(n) g' \left(\sum_{j=0}^{J} \beta_{kj} y_j \right) \beta_{kj} f \left(\sum_{i=0}^{I} \alpha_{ji} x_i \right) x_i. \tag{11}$$

Such weight adjustment equations can be calculated for a network with any number of layers. In practice, the networks in this research implemented a local gradient signal which was propagated back through the network during training. The local gradient of a node is the partial derivative of error energy for a particular observation, $E(n)$, with respect to the induced local field of a node. Haykin (1999) provides a concise presentation of local gradient calculation.

The temperature prediction models included several constant parameters selected from Smith et al. (2006). All networks had a single hidden layer consisting of 120 nodes that were distributed equally among the three slabs comprising the layer. Limited experimentation with the number of hidden nodes had only minimal effect on accuracy. The use of 120 nodes produced a slight improvement in accuracy. All initial weights were in the range [−0.1,0.1]. The learning rate, η, was set to 0.1 and no momentum term was used. Smith et al. (2006) established that comparing alternative ANN models by instantiating and training multiple networks for each model led to improved parameter selection and a reduction in prediction error when compared to previous air temperature prediction research (Jain et al., 2006). Because the EBP algorithm attempts to minimize prediction error by performing gradient descent in the space of possible weights, networks tend to converge towards local optima. Making use of repeated instantiations of the same model, differing only in the random seed used to generate initial weights and the order of training pattern presentation, is analogous to making repeated draws from the distribution of possible local optima.

2.3. Experiments

Each network created during the course of this research required between 6 and 20 h to train and evaluate using one of 30 Dell Pentium IV workstations in a University of Georgia computer laboratory. When evaluating the accuracy of competing ANN models, each alternative was instantiated 30 times for the four-hour prediction horizon. All model comparisons for parameter selection were made on the basis of the four-hour horizon and then implemented across all other horizons. While it was possible that other horizons might have benefited from alternative design decisions, the amount of processing time necessary to investigate all horizons was prohibitively large and beyond the scope of this study. Prior work had shown that selecting model parameters on the basis of the four-hour horizon reduced MAE across all horizons (Smith et al., 2006).

The most accurate network for each competing model, measured in terms of MAE over the development set, was used to represent the performance of that model. Once a representative instantiation was assigned to a model, it was evaluated over the selection set. When comparing among alternatives, the model with the lowest MAE over the selection set was chosen.

One aspect of this research involved comparing the accuracy of winter and year-round models. Networks instantiating the year-round model were trained over a data set of 300,000 patterns for 15 epochs, or 4.5 million learning events. For each instantiation, the year-round model randomly selected a different subset from the more than 1.25 million training patterns available. Year-round instantiations, therefore, differed from one another not only in the initial weights and the order of the training patterns, but also in the subset of training patterns itself. This restriction on the number of training patterns was introduced to allow for a fair comparison between the year-round and winter models. Winter networks made use of all of the approximately 300,000 patterns in the winter-specific subset of development patterns for training.

Jain et al. (2006) found that the inclusion of rainfall variables did not increase the accuracy of temperature prediction, while Ramyaa (2004) concluded that rainfall was helpful in classifying upcoming periods as containing a freeze, near-freeze, or non-freeze event. Both studies relied on single-network evaluation to select inputs. Smith et al. (2006) did not address this issue, but arbitrarily included rainfall data as inputs. An experiment was conducted herein to determine the effects of using rainfall variables as inputs to the year-round model. Thirty networks were trained over randomly selected, 300,000-observation subsets of the development set for 15 epochs. These ANNs instantiated models with and without rainfall variables as input.

Several additional experiments were conducted to determine the utility of changes to the input vector and the output range. A model was instantiated that included additional values and rates of change for observations 15, 30, 45, 75, 90, and 105 min prior to the point of prediction. Because of the additional temperature, solar radiation, wind speed, humidity, and rainfall variables, the model used an input vector with 318 values. Networks developed with these additional inputs did not provide more accurate predictions. Likewise, no improvement in accuracy was found in a model with the values and rates of change corresponding to 13, 15, 17, 19, 21, and 23 h prior to the prediction point removed from the input vector. A model was also developed with the hyperbolic tangent function used for the activation function of the output node. This change, which doubled the output range of the network from [0,1] to [−1,1], did not improve accuracy and was not investigated further.

Two ensemble techniques were also investigated. Series aggregation involves improving the output of a single model by training a second model using the outputs or errors of the first and combining their outputs. Series aggregation was implemented by training a network to predict the errors of the most accurate year-round ANN, as measured by development set MAE. Once again, a four-hour horizon was used to study this approach. Thirty instantiations of the four-hour series model were trained using randomly selected, 400,000-observation subsets of the development set for 10 epochs,

Table 1
Year-round model prediction accuracies over the selection and evaluation data sets.

Horizon length (h)	Mean absolute error (°C)	
	Selection data set, 2001–2003	Evaluation data set, 2004–2005
1	0.525	0.516
2	0.834	0.814
3	1.046	1.015
4	1.226	1.187
5	1.404	1.356
6	1.483	1.432
7	1.577	1.532
8	1.669	1.623
9	1.734	1.686
10	1.801	1.755
11	1.865	1.815
12	1.908	1.873

or 4 million learning events. Parallel aggregation involves the development of a meta-model that aggregates the outputs of several primary models. This approach was conducted by including the outputs of the five most accurate existing networks as additional inputs to a new ANN model. Thirty networks instantiating a four-hour parallel model were trained for eight epochs over the entire development set, approximately 10 million learning events per network. These approaches represent two methods by which previously developed ANNs that had settled into local optima could be improved.

3. Results and discussion

The first experiment compared the accuracy of year-round and winter models when evaluated on a data set consisting of the winter observations contained within the selection set described previously. Each model was instantiated by 30 distinct initial networks differing only in the initial random weights and the order of presentation of training patterns. Year-round networks were trained over a randomly chosen subset of 300,000 development patterns of the more than 1.25 million patterns available. Such a restriction was necessary for an accurate comparison to winter models. The representative network for any model was the instantiation that minimized MAE over its development set. The winter model differed from other models in that its development set was restricted to the winter subset of development patterns. Over the winter subset of the selection data set, the winter model MAE was 1.414 °C while the year-round model MAE was 1.416 °C. These comparable errors suggested that expanding the coverage of ANN models to the entire year would not sacrifice the accuracy of temperature prediction during the winter.

In the second experiment, the year-round model was compared to a similarly trained model with rainfall excluded as an input. Removing rainfall variables reduced the number of inputs from 258 to 208 and the number of weights from 31,210 to 25,201. The year-round model including rainfall as an input produced a selection data set MAE of 1.239 °C. The model without rainfall inputs produced a selection data set MAE of 1.259 °C. The no-rainfall model was also less accurate for the winter selection subset, generating an MAE of 1.428 °C compared to 1.416 °C. It is likely that the use of single-network evaluation by Jain et al. (2006) led to the exclusion of rainfall variables as inputs in that study.

Having shown that the year-round model with rainfall was a suitable replacement for the winter model, training was performed with no restrictions on the number of development patterns. The accuracy improvements of the year-round model were consolidated by training networks over the 1.25 million available development patterns for eight epochs. For the four-hour prediction horizon, 30 year-round networks were trained using the entire development set for approximately 10 million learning events. The selection data set MAE of the representative network was 1.226 °C, an improvement of 0.013 °C compared to the results of the first experiment. This representative network is hereafter referred to as the standard network.

In the first of two ensemble experiments conducted for the four-hour horizon, the existing network was boosted. A group of 30 networks were trained to predict the errors of the existing network. The series aggregation model made use of 259 inputs: the original 258 as well as the output of the standard network. By combining the resulting network's output with that of the standard network, accuracy would be improved to the extent that the new network was successful in predicting the standard network's prediction errors. Following 10 epochs of training over a randomly selected subset of 400,000 development patterns (4 million learning events), the selection MAE decreased by less than 0.004 °C relative to the standard network. Though errors did decrease slightly, the reduction in error was negligible while the training was time-consuming. As such, series aggregation was not found to be a useful method of improving the prediction accuracy for this problem.

An experiment to explore parallel aggregation was implemented by training a four-hour prediction model with 263 inputs: the 258 provided to the year-round model as well as the output of the five most accurate year-round networks trained over the complete development set. After training was complete, the networks had undergone more than 10 million learning events. The 1.220 °C selection MAE of the representative network for the parallel model improved upon the four-hour standard network by approximately 0.006 °C. These improvements did not merit the additional computational time required to implement the model across all horizons.

Table 2
Comparison of model prediction accuracies over the winter selection and evaluation subset.

Horizon length (h)	Winter selection 2001–2003 mean absolute error (°C)		Winter evaluation 2004–2005 mean absolute error (°C)	
	Winter model[a]	Year-round model	Winter model[a]	Year-round model
1	0.534	0.529	0.527	0.522
2	0.884	0.883	0.864	0.860
3	1.167	1.167	1.118	1.117
4	1.401	1.398	1.338	1.331
5	1.624	1.615	1.546	1.535
6	1.811	1.793	1.715	1.695
7	1.987	1.930	1.874	1.831
8	2.126	2.081	2.007	1.958
9	2.243	2.213	2.091	2.055
10	2.362	2.311	2.191	2.161
11	2.443	2.417	2.250	2.239
12	2.526	2.495	2.333	2.299

[a] Smith et al. (2006).

The performance of the ensemble techniques suggested that the networks were not settling into easily improvable local optima.

Models were then developed for the other prediction horizons from 1 to 12 h ahead using inputs determined from the four-hour standard model. These models made use of 258 inputs, including air temperature, wind speed, relative humidity, solar radiation, and rainfall and hourly rates of change at the time of prediction as well as the history of prior observations at 1-h intervals going back 24 h. Also among the models' inputs were four cyclic time-of-day and four cyclic day-of-year terms. For each model, 30 networks were trained over all available development patterns for eight epochs. Due to the computational time required to train and evaluate so many networks over such a large number of patterns, poorly performing networks were discarded at two stages of the training process. Following two epochs of training (approximately 2.5 million learning events), each network was evaluated over a 100,000-pattern subset of development patterns and the 15 networks with the highest MAEs were discarded. Following another two epochs, the process was repeated and five of the remaining 15 networks were discarded. The remaining networks were trained for an additional four epochs, for a total of more than 10 million learning events each.

The MAEs of the standard models at all prediction horizons for both the selection and evaluation sets are presented in Table 1. The MAE of the selection data set increased monotonically from 0.525 °C at the one-hour horizon to 1.908 °C at the twelve-hour horizon. The trend was similar for the MAEs of the evaluation data set that were calculated for the same sites during 2004 and 2005. The MAEs of the evaluation data set associated with the models during this period were slightly lower than those for the selection set, ranging from 0.516 °C for the one-hour model to 1.873 °C for the 12 h model. The four-hour horizon that was used as the basis for experimentation in this research resulted in an MAE of 1.187 °C for the evaluation data set compared to 1.226 °C for the selection data set.

The resulting year-round models, developed with all 1.25 million available patterns, were also evaluated over the winter selection and winter evaluation subsets (Table 2). This allowed for further comparison with the winter models previously developed by Smith et al. (2006). When compared to these winter models, the year-round ANN errors were less than or equal to the winter selection and winter evaluation MAEs for all horizons. These results indicated that the year-round air temperature prediction models, developed to be included in general decision support aids, were suitable for use during the winter months. When comparing the results presented in Tables 1 and 2, the MAEs for the winter data set proved to be higher than those for the year-round data across all horizons.

The scatter plots depicted in Fig. 2 present the prediction errors for four horizons. As expected, differences between predicted and observed temperatures increased with the prediction horizon as reflected in the greater dispersion about the 1:1 line of a hypothetical, perfect model. Likewise, as horizon length increased, the value for R^2 decreased. The regression analysis also suggested that the models for longer horizons had a tendency to over-predict low-temperature observations and under-predict high-temperature observations.

The accuracy of the models was also evaluated in relation to the day of year and time of day at three different prediction horizons: 4, 8, and 12 h. Fig. 3 presents a contour plot of MAE for the combined selection and evaluation sets of 2001–2005 partitioned by week of year and hour of day. The data sets were combined to

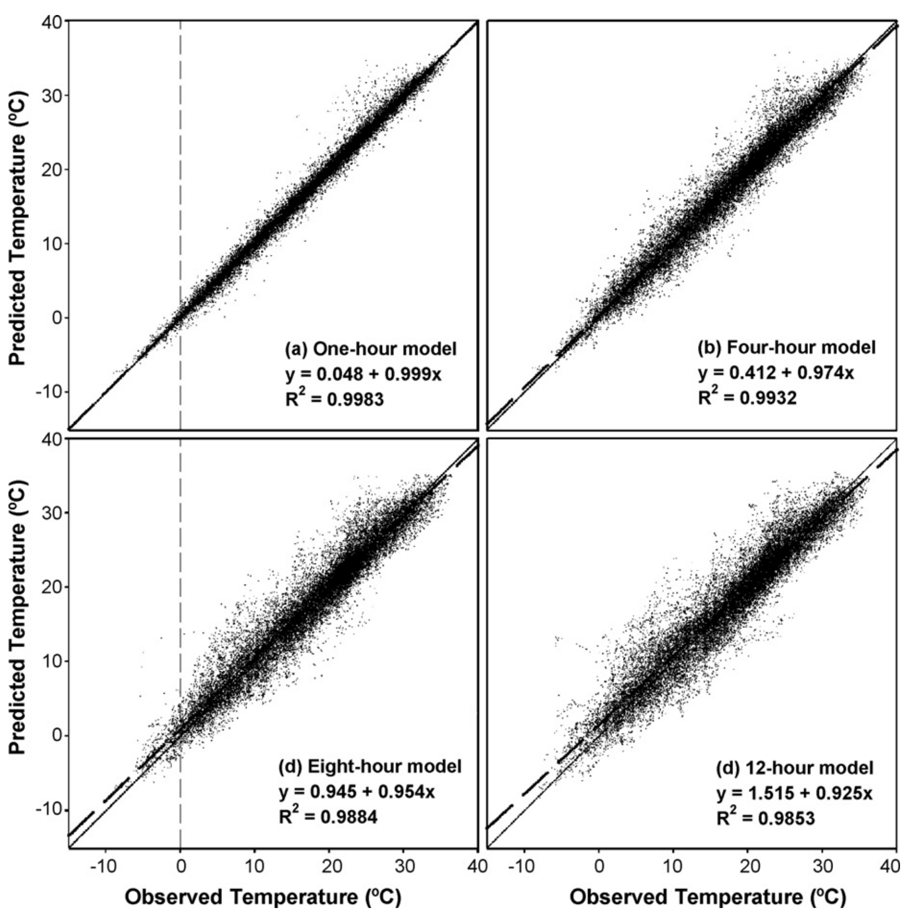

Fig. 2. Predicted and observed temperatures for Byron during 2004–2005 for the standard model at the (a) one-hour horizon, (b) four-hour horizon, (c) eight-hour horizon, and (d) twelve-hour horizon. A solid diagonal line indicates a hypothetical perfect model.

reduce the impact of weather trends from any single year. Each point corresponds to the average of all MAEs associated with a particular week of the year and hour of the day. For example, the MAE during the third week at 1200 h was calculated over all predictions made for 1200, 1215, 1230, and 1245 h during January 15–21 for the years 2001–2005 at each of the 13 sites in the combined data set. The predictions themselves were based on data available at a point 1–12 h prior to the observation time, depending on the prediction horizon. At the four-hour horizon (Fig. 3a), the standard model was most accurate when predicting overnight/early-morning temperatures, especially during the summer. Most of the MAEs for summer nights and early mornings were less than 1 °C. The prediction MAEs were highest for the four-hour model between 0900 and 1445 h from fall through early spring (weeks 46 through 12) and between 1400 and 1945 h during the summer (weeks 22 through 35).

For the eight- and twelve-hour horizons shown in Fig. 3b and c, the magnitudes of the prediction errors during the summer nights and mornings were also smaller, relative to the rest of the day. During the summer, the eight- and twelve-hour horizons also showed

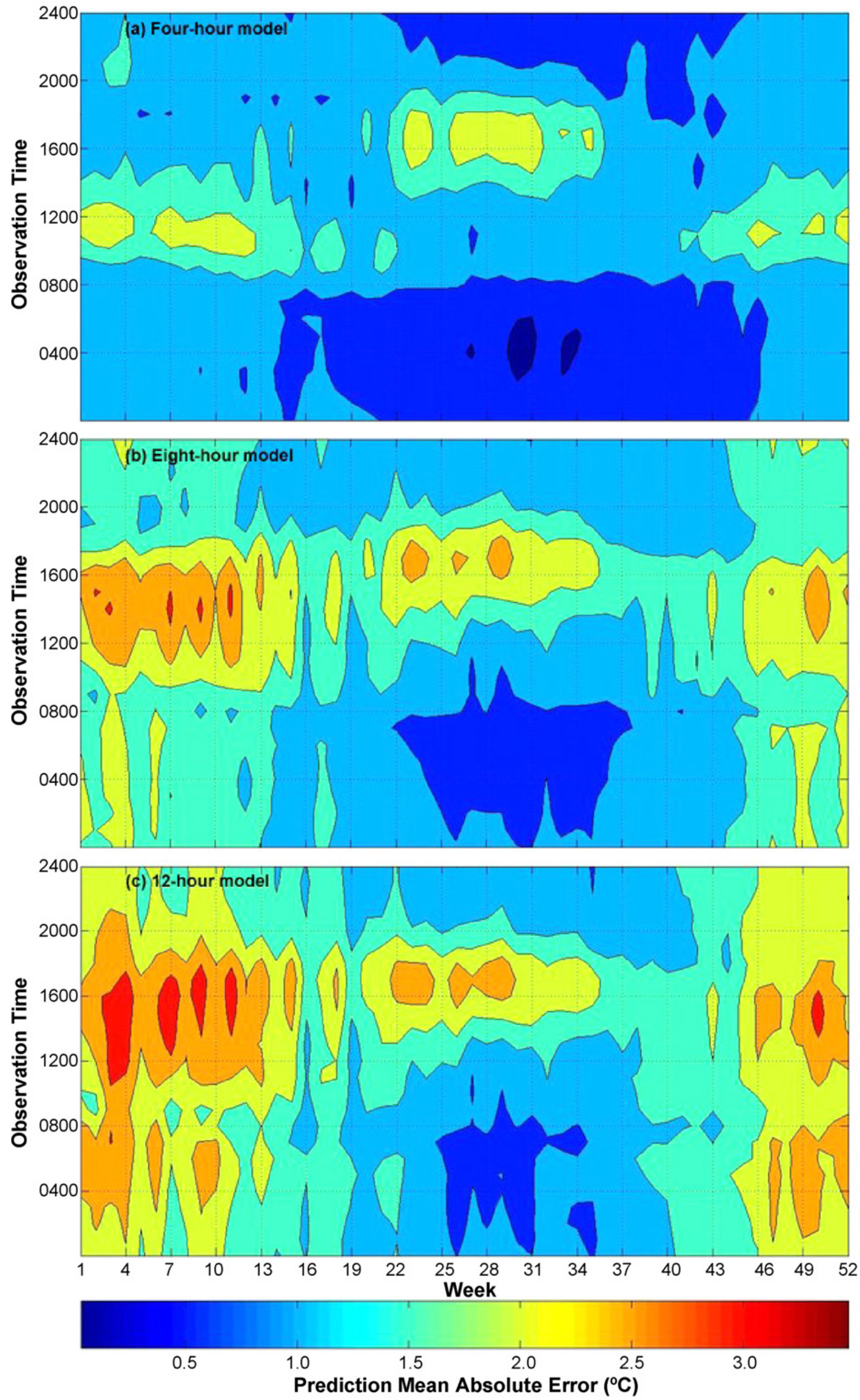

Fig. 3. Prediction MAE across all evaluation sites during 2001–2005, partitioned by week and hour. The observation time indicates the time for which predictions were made.

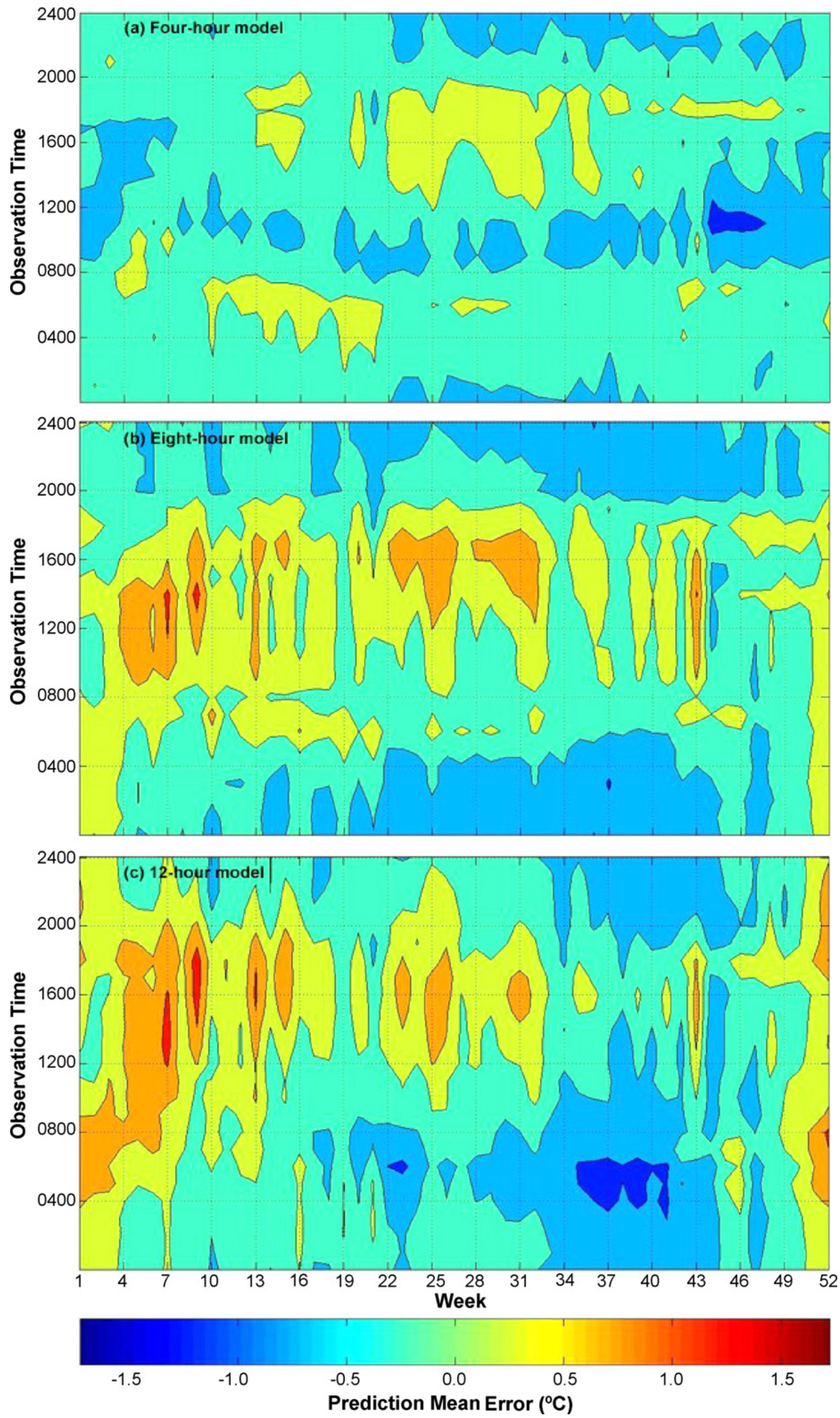

Fig. 4. Prediction mean errors across all evaluation sites during 2001–2005, partitioned by week and hour. The observation time indicates the time for which predictions were made.

the highest MAEs during the afternoons from approximately 1400 to 1900 h. The behavior of the standard eight- and twelve-hour model differed substantially from the four-hour horizon during the rest of the year. The eight- and twelve-hour horizons had noticeably larger errors during the fall and winter mornings than the four-hour horizon. The errors for the twelve-hour horizon were larger still. Periods of higher-than-average MAE for the four-hour horizon occurred in the morning and early afternoon outside of the summer. At the eight- and twelve-hour horizons, these periods of lower accuracy persisted until the early evening.

The bias of the standard models as measured by mean error partitioned by week and hour is presented in Fig. 4. A negative mean error indicates a tendency to under-predict, while a positive mean error is evidence of over-prediction. Fig. 4a shows that the two areas of greatest MAEs identified at the four-hour horizon, fall/winter mid-day and summer afternoon, were associated

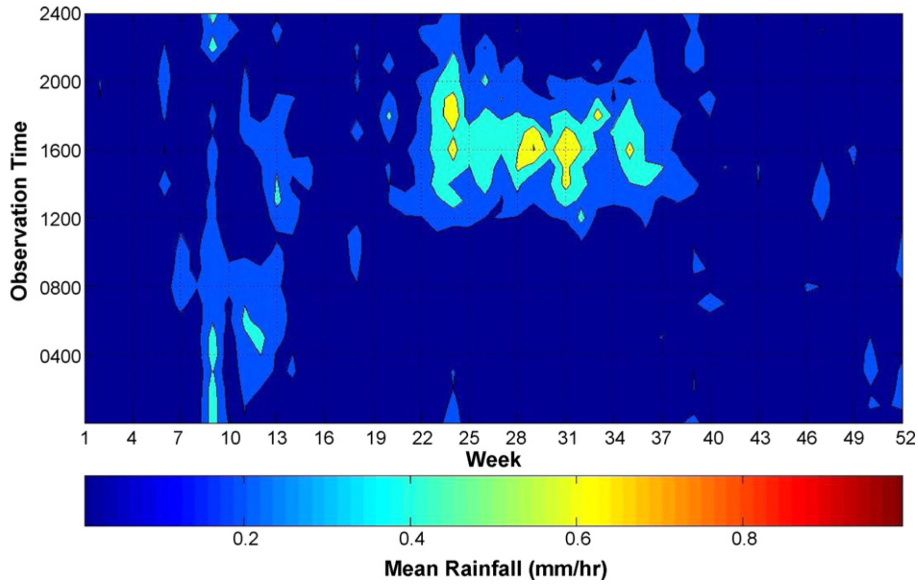

Fig. 5. Mean rainfall across all evaluation sites during 2001–2005 partitioned by week and hour.

with at least two distinct phenomena. During the fall and winter at midday, the largest errors showed a negative bias, suggesting that the unanticipated weather events were warming events. During the summer afternoons, bias was positive, suggesting the presence of unpredicted cooling events. For the eight- and twelve-hour horizons (Fig. 4b and c) the summer afternoons were also associated with a positive bias and periods of maximum bias were closely associated with high MAEs. Biases associated with winter errors for each of these horizons were also largely positive.

The unanticipated cooling events that occurred during the summer afternoons across all horizons are likely associated with rain showers and thunderstorms. Fig. 5 presents mean daily rainfall partitioned by week and hour for the same sites and periods. Summer afternoons and evenings clearly accounted for the most active periods of rainfall. Moreover, these times corresponded to the weeks and hours associated with relatively higher MAEs and biases for the four-, eight-, and twelve-hour horizons. This strongly suggests that the unanticipated summer cooling events were associated with rainfall or the corresponding cloud cover.

The air temperature predictions provided on the AEMN website are presented to users as a sequence of twelve temperatures generated at the same time, each corresponding to a prediction horizon of 1–12 h. Two such sequences are presented alongside observed air temperatures in Fig. 6 for 1 and 2 March 2005 for Dearing. A sustained freeze began at 2300 h on March 1 and lasted until 0830 h the following day. A minimum temperature of −2.564 °C occurred at 0700 on 2 March. The first of the two prediction sequences was generated at 1800 h on 1 March, 5 h before the first freezing temperature was recorded. This sequence predicted the time of onset accurately and had absolute errors that were less than 1 °C for the first 10 h of prediction. Such a sequence of predictions would provide time for fruit growers who use the AEMN website to obtain temperature predictions for their local area and take steps to mitigate crop damage. The second prediction sequence was generated at midnight, 1 h after the beginning of the freeze, and was consistent with the first. In addition, the second track accurately predicted the rate of warming following the minimum overnight temperature.

The observed air temperatures and prediction sequences for 28 and 29 July 2005 at Homerville are displayed in Fig. 7. During this period, high temperatures caused heat exhaustion in several farmers and reduced pesticide efficacy in the south-central Georgia Agricultural Statistical District that includes Homerville (NASS, 2005). On 28 July in Homerville, temperatures climbed above 35 °C with two periods of cooling during the afternoon, the first between 1445 and 1545 h and a second, more rapid cooling event beginning at 1645 h. Both of these cooling periods coincided with a brief decrease in solar radiation not typical for clear-sky conditions. The cause of this reduction in solar radiation was presumably cloud cover. A brief rainfall of more than 5 mm occurred between 2130 and 2200 h on the night of the 28th. This was associated with a third, but smaller, unanticipated cooling during this time. The first prediction sequence, generated at 0600 h on the 28th, accurately predicted observed temperatures within ±1.1 °C for all but the final prediction at 1800 h. The second sequence was generated at noon and also closely predicted observed temperatures. Prediction errors were within ±0.5 °C for the five temperature predictions prior to 1800 h. Following the second cooling event, the prediction error at 1800 h was more than 3.4 °C, with subsequent, late-night observed air temperatures more closely reflecting the 1200 predic-

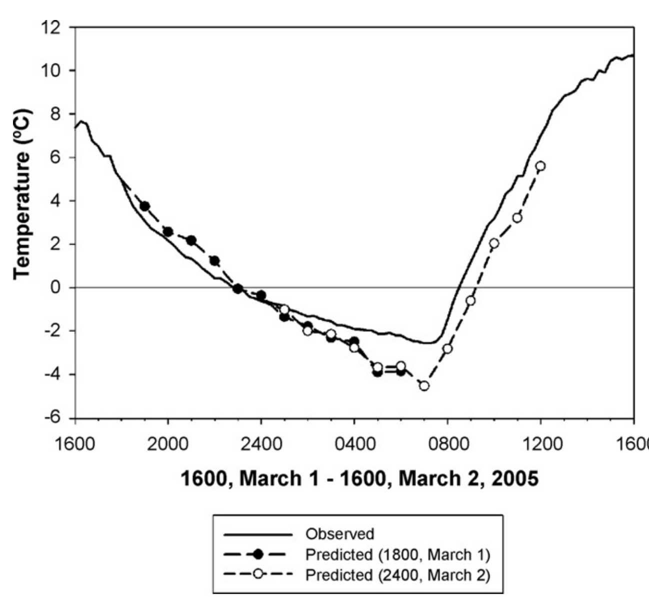

Fig. 6. Observed and twelve-hour prediction sequences of temperature during 1–2 March 2005 for Dearing, GA.

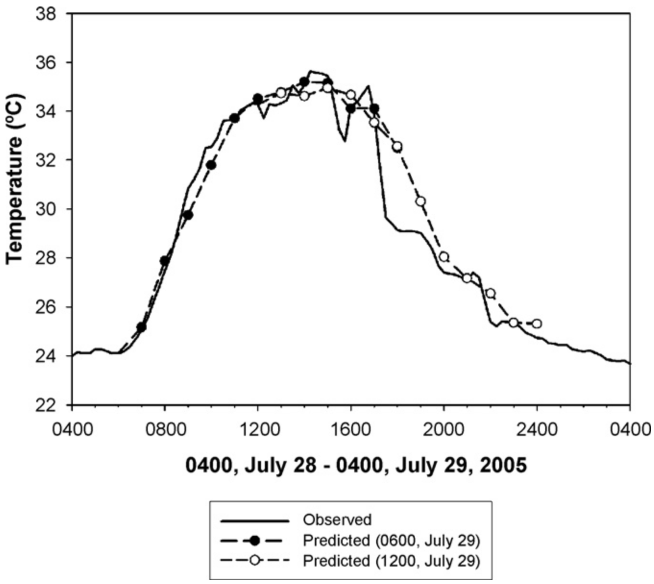

Fig. 7. Observed and twelve-hour prediction sequences of temperature during 28–29 July 2005 for Homerville, GA.

tion sequence. Evaluations of multiple sites over a five-year period, coupled with the examination of individual prediction tracks, indicated that unanticipated cloud cover and rainfall were the likely cause of prediction errors associated with cooling events.

4. Summary and conclusions

Year-round air temperature prediction models were developed for prediction horizons of 1 to 12 h using Ward-style ANNs. These models were intended for use in general decision support and are currently implemented on the AEMN website, www.GeorgiaWeather.net. The implementation of these models is a clear example of research-based Artificial Intelligence (AI) that has been implemented in a practical manner in a real-world application (Farkas, 2003). The ANN design modifications described herein provided increased accuracy over previously developed, winter-specific models during the winter period. It was shown that models that included rainfall terms in the input vector were more accurate than those that did not. Applying the ensemble techniques of series and parallel aggregation to single-network models was not found to be useful for this problem domain, as the very modest improvements in prediction accuracies came with a heavy computational cost. The accuracy of the resulting four-, eight-, and twelve-hour models was analyzed, showing that unanticipated cooling events were the most significant obstacle faced, especially at longer horizons.

The results suggest that accurate cloud-cover predictions might aid in the prediction of associated cooling events, especially during the summer. Further study might also determine if models specifically tailored to the periods of greater-than-average prediction errors might be useful in an ensemble approach. Because of the focus on developing models applicable to a broad range of locations, temperature prediction work in the AEMN domain has not made use of geographic information. Future work could focus on the possibility of adding the information in such a manner as to preserve the general applicability of the model.

Acknowledgments

This work was funded in part by a partnership between the USDA-Federal Crop Insurance Corporation through the Risk Management Agency and the University of Georgia and by state and federal funds allocated to Georgia Agricultural Experiment Stations Hatch projects GEO00877 and GEO01654. The authors would also like to thank the other members of the AEMN research team, Daniel Shank and Kevin Crowell, for their input and suggestions.

References

Bruton, J.M., McClendon, R.W., Hoogenboom, G., 2000. Estimating daily pan evaporation with artificial neural networks. Trans. ASAE 43 (2), 491–496.
Elizondo, D., Hoogenboom, G., McClendon, R.W., 1994. Development of a neural network to predict daily solar radiation. Agric. Forest Meteorol. 71, 115–132.
Farkas, I., 2003. Artificial intelligence in agriculture. Comput. Electron. Agric. 40, 1–3, Editorial.
Ghielmi, L., Eccel, E., 2006. Descriptive models and artificial neural networks for spring frost prediction in an agricultural mountain area. Comput. Electron. Agric. 54, 101–114.
Haykin, S., 1999. Neural Networks: A Comprehensive Foundation, 2nd ed. Prentice-Hall, Upper Saddle River, NJ.
Hoogenboom, G., 2000. The Georgia automated environmental monitoring network. In: Preprints of the 24th Conference on Agricultural and Forest Meteorology, pp. 24–25.
Jain, A., McClendon, R.W., Hoogenboom, G., 2006. Freeze prediction for specific locations using artificial neural networks. Trans. ASAE 49 (6), 1955–1962.
Jain, A., McClendon, R.W., Hoogenboom, G., Ramyaa, R., 2003. Prediction of frost for fruit protection using artificial neural networks. American Society of Agricultural Engineers, St. Joseph, MI, ASAE Paper 03-3075.
Jiang, D., Zhang, Y., Hu, X., Zeng, Y., Tan, J., Shao, D., 2004. Progress in developing an ANN model for air pollution index forecast. Atmos. Environ. 38 (40), 7055–7064.
Li, B., McClendon, R.W., Hoogenboom, G., 2004. Spatial interpolation of weather variables for single locations using artificial neural networks. Trans. ASAE 47 (2), 629–637.
Maqsood, I., Khan, M.R., Abraham, A., 2004. An ensemble of neural networks for weather forecasting. Neural Comput. Appl. 13, 112–122.
National Agricultural Statistical Service, 2005. Crop Progress and Condition Report. 31 July 2005.
Prabha, T., Hoogenboom, G., 2008. Evaluation of the weather research and forecasting model for two frost events. Comput. Electron. Agric. 64, 234–247.
Ramyaa, 2004. Frost prediction using artificial neural networks: a classification approach. MS Thesis, Artificial Intelligence Center, University of Georgia, Athens, GA.
Ruano, A.E., Crispim, E.M., Conceição, E.Z.E., Lúcio, M.M.J.R., 2006. Prediction of building's temperature using neural networks models. Energy Build. 38, 682–694.
Shank, D., Hoogenboom, G., McClendon, R.W., 2008a. Dew point temperature prediction using artificial neural networks. J. Appl. Meteorol. Climatol. 47 (6), 1757–1769.
Shank, D.B., McClendon, R.W., Paz, J.O., Hoogenboom, G., 2008b. Ensemble artificial neural networks for prediction of dew point temperature. Appl. Artif. Intell. 22 (6), 523–542.
Smith, B.A., McClendon, R.W., Hoogenboom, G., 2006. Improving air temperature prediction with artificial neural networks. Int. J. Comput. Intell. 3 (3), 179–186.
Ward Systems Group, Inc. NeuroShell 2, Frederick, MD (http://www.wardsystems.com/).
Yi, J., Prybutok, R., 1996. A neural network model forecasting for prediction of daily maximum ozone concentration in an industrialized urban area. Environ. Pollut. 92 (3), 349–357.